Clarifications & Response to Issues Raised in:

"Estimating Hospital Deaths Due to Medical Errors: Preventability Is in the Eye of the Reviewer"
Rod Hayward & Tim Hofer
JAMA, July 25, 2001 - Vol 286, No. 4

Our article published in the July 25, 2001 issue of JAMA regarding estimating the impact of medical errors on hospital mortality has generated much discussion. We have been pleased with the very positive response to the paper. There has been some excellent dialogue and some expected skepticism but much of the skepticism has been fueled by some misinterpretations and misinformation. Below you will find three documents that we hope will help clarify important aspects of this paper: 1) A more detailed response to a letter by P. Barach that is to be published in JAMA's letters to the editor, 2) Clarifications and responses to issues raised by others in other forums, and 3) an Appendix to the paper that includes some information on the study's methods that could not be fit into the published article.

Expanded Response to a Letter to the Editor by P. Barach

To the Editor:(1)

The Harvard Medical Practice Study (2) is one of the most important studies to be conducted in the past 20 years and produced vital information on the incidence and consequences of adverse events. However, it was primarily designed to estimate the number of adverse events, not the preventability of adverse events, and our understanding of important issues in the reliability of peer review of quality were very preliminary at the time the HMPS was conducted. The HMPS authors were therefore appropriately cautious about the statistics on preventable adverse events in their publications. Almost all previous studies estimating the preventability of adverse events using implicit peer review, including the HMPS, have assumed: (1) the error assessment had unbiased inter-rater reliability and (2) if care had been optimal the patient had a 100% chance of not dying. Our study clearly found that both assumptions are incorrect and lead to dramatic over-estimations.

We see no merit in Barach's concern that our study is too small. The 179 deaths reviewed (including 111 active-care deaths) are comparable in number to the HMPS, (2) represent all deaths occurring in over 5,000 admissions and the 95% CIs for our main findings are 1.0%-1.5% and 0.3%-0.7%. (3) Therefore, unless knowing precisely where the true value lies

between 1.0% and 1.5% is important for some reason unclear to us, our 383 reviews of 179 deaths is more than adequate.

Barach was very concerned about generalizability. We feel that such concerns are certainly justified when considering the third issue addressed in our paper, the probability of living 3 months or longer with good cognitive function, and we were careful to caution the reader about this limitation in our paper. However, the first two sources of over-estimation (which accounted for the majority of the overestimation [see above]) are intrinsic to the analytic approach used in previous studies (not to the generalizability of the study population) and are therefore almost certainly applicable to previous studies, including the HMPS statistics. We have requested the HMPS data or key tables so that we can demonstrate this directly with HMPS data. Our study suggests that substantial over-estimation will occur in any study that estimated the preventability of adverse events using these same invalid assumptions.

Barach's concern about the over-sampling of cases with laboratory abnormalities is without merit since, as we clearly state in the paper, we used sampling weights to account for this over-sampling. Even if for unclear reasons he is concerned that sample weights are inadequate for this purpose, the over-sampled cases had higher preventability ratings, so the over-sampling could only have inflated our preventability estimates. He doesn't specify, and it is unclear to us, how he believes that standardizing peer assessment on a severity index would have improved our results. Finally, Barach erred when he computed that 45% of cases were deemed to be suboptimal or worse as he inadvertently double counted some numbers and then added together categories that are not mutually exclusive (actual frequency=26%). More importantly, he attempts to quote the same naïve statistics that past studies have quoted and ignores our main finding - that these statistics are invalid. Unless you adjust for the skew in inter-rater reliability, you greatly over-estimate how the median reviewer would rate these cases.

Finally, we completely agree with Barach that there are major quality and patient safety problems in our hospitals and that we should be less worried about the numbers and be more concerned about providing better and safer care. We can and must do better. In fact, we believe that a strong case can be made that a comprehensive quality improvement and patient safety program could save 100,000 lives a year, but only if we devote more time and attention to outpatient quality, target common and high-impact quality problems and work harder to understand the nature of the problem and rigorously test proposed quality enhancement and safety interventions. We agree that we cannot improve quality if we do not maintain and enhance the quality of health care teams and adapt systems to the unique needs and circumstances of the specific setting where healthcare is being delivered. However, we also feel that we are not likely to enhance quality by just leaving these teams alone to work on safety and quality. Coordinated programs to facilitate accountability, information sharing and feedback through reporting, quality monitoring and rigorous evaluation of proposed quality improvement interventions are essential. The purpose of reporting and quality monitoring should not be directed at "counting the numbers", but should help us better understand and solve the problem. Specious statistics are not helpful in this effort. If we are to improve healthcare quality, we need more rigorous attention to the nature and magnitude of specific quality and specific patient safety problems and need to

aggressively pursue problems in proportion to their seriousness and preventability. Sensationalized and misleading statistics just result in misdirection and confusion, and can divert us from pursuing, and achieving, the greater public good.

Rodney A. Hayward, MD
Director, VA Center for Practice Management & Outcomes Research
VA Ann Arbor Healthcare System
Professor of Medicine & Public Health
University of Michigan

Timothy P. Hofer, MD, MSc
Senior Investigator, VA Center for Practice Management
& Outcomes Research
VA Ann Arbor Healthcare System
Associate Professor of Internal Medicine
University of Michigan School of Medicine

References:

1. This document can be found at .

2. Brennan TA, Leape LL, Laird NM, Herbert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med 1991 Feb;324(6):370-6.

3. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. JAMA 2001;286:415-420.

4. Hofer TP, Kerr EA, Hayward RA. What is an error? Eff Clin Pract 2000;3(6):1-10.

Clarifications & Responses to Issues Raised Regarding:

"Estimating Hospital Deaths Due to Medical Errors: Preventability Is in the Eye of the Reviewer"
Rod Hayward & Tim Hofer
JAMA, July 25, 2001 - Vol 286, No. 4

1) We are supportive of the Patient Safety movement.

We regret that some have perceived our paper as a challenge to the importance of the Patient Safety movement, but our paper clearly states that we believe patient safety is an important problem. We support the general message of the IOM reports and are glad that patient safety is finally being taken seriously. We believe that patient safety is an essential part of any responsible health care delivery system and certainly do not feel that our paper should be interpreted by anyone as detracting from it's importance. Patient safety is not about the numbers. It is about paying attention to the environment where care is delivered so as to do the most good with the least harm. We feel that every conscientious provider should support and assist in these efforts to enhance quality and patient safety.

However, we do not believe that "supporting patient safety" means accepting claims and statistics without question. We feel that research on the nature of the problem should be scientifically rigorous and presented accurately and that the data and statistics should not be sensationalized for short-term political gains. We think that it is important to attempt to understand correctly the data and statistics if we are to improve upon the problems with which we are faced. Misleading and exaggerated statistics can be counterproductive. Past statistics have often been used to make it sound like hospitals are very dangerous places that should be avoided at all cost, however, we believe that at closer look at the actual data suggests that hospital are doing more good than ever before, but we are missing many opportunities to prevent even more deaths and that we can do so with even less harm. Although most of the attention has been directed to our study's finding that past studies have over-estimated the number of preventable deaths, we feel that the more important finding in the paper is that past studies have misunderstood the nature of "errors" and "quality problems" that are found in implicit chart review (see # 2 below).

2) Some were concerned that the differences between our study and the Harvard Medical Practice Study (HMPS) statistics were mainly due to our study considering 3-month prognosis and some have questioned the ethics of considering this factor.

We did consider 3-month prognosis, but this was only one of three factors we found to be important in estimating the impact of hospital quality problems on mortality and it was clearly not the most important factor. If 3-month prognosis had been the only or even the main finding, we would not have sought publication in JAMA and we suspect that JAMA would not have accepted the paper if we had submitted it to them. Instead, our study is the first study to examine what it means when a reviewer says that a death is definitely or probably preventable or due to error. The HMPS and Utah/Colo study are two of the most important studies in the last 10-15 years, however, neither was designed for the purpose of estimating the number of preventable deaths and the authors clearly warned people that these statistics might be substantial over-estimates. (2, 13) These statistics were calculated with an underlying assumption that when one or two reviewers rate a death as probably or definitely due to error that (1) the error assessment was either highly-reliable or had unbiased inter-rater reliability and that (2) if care had been optimal the patient had a 100% chance of not dying. Our study clearly demonstrates that both assumptions are incorrect and lead to substantial over-estimations. Table 3 in the paper shows the individual impact of each of the 3 sources of over estimation; we tried to emphasize this point in the paper (see page 419, column 1, paragraph 1; page 419, column 2, paragraph 1; and page 419, column 3, the last paragraph.)

The first source of over-estimation can be conceptualized as past statistics not considering

the attributable risk of the perceived quality problem. Assuming that optimal care would always prevent a death would be reasonable if the types of quality problems found were egregious, highly fatal errors, such as inadvertently giving toxic doses of medication. However, no such egregious fatal errors were found either in our study or in the HMPS. Indeed, one of the prime examples given of a major fatal error in the HMPS was a patient with atrial fib who was not acutely anticoagulated and had a stroke during his hospital course.3 This example hardly qualifies as the sort of clearly fatal error on which one could base these assumptions. First, we are aware of no evidence establishing a benefit of emergent anti-coagulation for a. fib. Second, even long-term anti-coagulation of a. fib only has a 10%-30% relative risk reduction (RRR) for mortality. This is not an unusual occurrence in medicine since few treatments in all of medicine have a mortality RRR greater than 30%. Therefore on average the previous approach to estimating the impact of errors on mortality was found to over-estimate the reviewers' rating of preventability by 2 to 3 fold (page 418, column 2). There is every reason to believe that the HMPS statistics also suffer from this source of true over-estimation and that this source of over-estimation is substantial. Taken literally, the method used is akin to estimating the public health impact of achieving optimal use of early revascularization for AMI by assuming that if all eligible patients receive optimal treatment that it would eliminate all hospital death in this group.

However, another source of over-estimation is even more important, both in terms of the magnitude of over-estimation and in terms of its significance in understanding the nature of the quality problems identified in past studies. We did not anticipate this finding and feel that it is perfectly understandable that past researchers missed it. They certainly did not have the data necessary to examine this source of over-estimation and this source of over-estimation was not predicted by those who criticized other aspects of the estimates.(2, 4-5) As we say in the paper (page 419, column 1),

"*Previously, most have framed ratings of preventable deaths as a phenomenon in which a small but palpable number of deaths have bad errors that are being reliably rated as causing death. Our results suggest that this view is incorrect. Our study suggests that if you have many reviewers, almost all deaths will have some reviewers who strongly feel that death could have been avoided by different care, however, most of the "errors" identified in implicit chart review appear to represent outlier opinions in cases in which the median reviewer either feels that an error didn't occur or had little or no effect on the outcome.*"

We believe this to be the most important point in our paper. The reason the inter-rater reliability of implicit review is so poor (usually between 0.2 and 0.4 for quality and error assessments) is not the fault of the method or the reviewers, it's an inevitable result of the fact that almost every hospital admission involves very difficult decisions in which you will get many opinions, often very strong opinions, if you ask many physicians. However, most of these questionable decisions involve aspects of medicine in which there is little evidence to help discern who's right. Since ratings of preventability are highly skewed, not accounting for the poor reliability of the ratings results in dramatically over-weighting outlier opinions. A careful examination of Tables 2 and 3 is important for understanding this point. This is a fundamental shift in our understanding of the nature of the problem.

Without asking each reviewer about the likelihood that the error actually caused the death, it was not possible for the HMPS and the Utah/Colo investigators to consider this source of over-estimation. However, we think that it is extremely likely that the HMSP found similar types of errors and therefore has a similar source of over-estimation. When an investigator from the HMPS gave three examples of errors from that study,3 he listed two examples of delays in diagnosis and one regarding lack of anti-coagulation for atrial fibrillation (which was much more controversial before 1990 and to our knowledge has still not been demonstrated to be beneficial in the acute hospital setting where instability may make temporarily postponing intervention a reasonable decision). If the typical "errors" in previous studies were more clear-cut blunders than in our study (ie, instances in which someone inadvertently received 10 times the recommended dose of chemotherapy") then the inter-rater reliability would be much higher than the 0.24 found in the HMPS (Note: our IRR was somewhat better than previous studies, suggesting that the previous statistics may be even greater over-estimates). There is every reason to believe that these first two sources of over-estimating the impact of quality problems found in our study are similar in the HMPS.

Let us now turn to what appears to be the most controversial aspect of our paper. Yes, we also considered short-term prognosis. We were as clear in the paper as possible to point out that this inclusion was meant to help evaluate the public health impact of the quality problems and does not make an error excusable or unimportant. We were also clear that this caveat might not be as generalizable to other populations as the first two issues our paper addresses. We have been criticize for considering this factor and some have termed these analyses as "offensive" and "unethical". We strongly disagree with this characterization and feel that this provided useful information and contradicts what many have asserted in the literature. Most importantly, we feel that reporting on potential life-years impacted should not be considered a radical, controversial or unethical concept in scientific discourse; it is just information. It should not necessarily be the major factor in deciding policy, but it should at least be considered when weighing the relative priority of competing quality and patient safety initiatives and policy options. Yes, if we had unlimited time and resources, then it becomes unimportant, but in the real world we feel that it is more ethical to examine the relative amount of public health benefit of competing problems and policy options than it is to ignore such factors.

3. Several critics have been concerned that our study is too small, whereas the HMPS and Utah/Colo studies reviewed thousands of hospital records.

For estimating deaths due to errors our study has statistical power that is comparable to that of the HMPS and is definitely superior to that of the Utah/Colo study. We should probably have made clearer in the paper that our study was also a very large study in which thousands of cases with adverse events were reviewed. For the purpose of estimating preventable deaths, however, it seemed to us that the number of deaths reviewed was more relevant. The 179 deaths reviewed (111 of them active-care deaths) represent all deaths occurring in over 5,000 admissions. Although we cannot find in the published literature the exact number of deaths reviewed in the HMPS, we estimate that the HMPS reviewed about 175 deaths for evidence of negligence. In addition, the greater number of independent chart reviews and the more precise estimates of attributable risk obtained on all reviews in our study gives us

comparable statistical power as the HMPS for estimating deaths due to errors. As a side-note, the Utah/Colo study included many fewer deaths and reviews of deaths than did our study (confirmed by a personal communication with Troy Brennan and the Utah/Colorado Study research team on July 31, 2001).

However, what is most important is not whether our study has as much statistical power as the HMPS in examining the impact of hospital quality on preventable deaths, but whether our study has adequate statistical power. One of the laws of statistical power is that it is much easier to get precise estimates of rare outcomes (such as a 1% to 2% incidence) than it is for outcomes with a moderate incidence (such as a 20% to 80% incidence). For estimating the incidence of rare outcomes you only need a very large sample if you are interested in a very precise estimate. The 95% CIs for our main findings are 1.0% to 1.5% and 0.3% to 0.7%. Therefore, unless it is felt that a precise measurement of exactly where the true value lies between 1.0 and 1.5 is important for some unstated (and to us unclear) clinical or public policy reason, we do not understand those who argue that our review of 179 deaths, which included 383 reviews of 111 active care deaths, is too small.

However, this is a methodological issue that could probably be easily resolved, and we make an open offer to any investigator who has assessed the impact of errors on adverse events to collaborate in the interests of clarifying this point. If we were to receive the cross-tabulation of the inter-rater reliability of two or more blinded independent reviews in such a study, we can explore the impact of our study's findings on their preventable adverse event estimates. We think that it is extremely likely that the first two sources of over-estimation will be similar in these other studies, since these sources of over-estimation are intrinsic to the analytic and review process used, not to the generalizability of our patient population. Since the over-estimation is due to the analytic approach and our study had similar initial findings and inter-rater reliability, we do not expect a major difference due to the VA sample, or due to our reviewers being "different" or due to temporal trends. Our study suggests that substantial over-estimation will occur in any study that estimated the preventability of adverse events without considering the attributable risk of the quality problem and the skew in inter-rater reliability.

4) Some have been concerned that it was inappropriate to extrapolate from the VA study to make national estimates.

First, we want to be clear that such extrapolations were made by the media and not by us. We believe that people have been too obsessed with the total number and not with parts of the paper that discuss the nature of the "errors". The press highlights, "It's really only 5,000 to 15,000 preventable deaths", although these numbers are not mentioned anywhere in our article (they are obtained by multiplying the preventability rates in our paper, with and without considering the 3-month prognosis estimate, by the number of admissions in the country, something that we did not feel was appropriate).

However, the most important reason that we feel it's inappropriate to extrapolate our results to make statements like, "It's really only 5,000 to 15,000 preventable deaths" is not because

the study was done in the VA, since 2 of the 3 phenomena found in our study should be highly transportable (see #2 above). Rather, it is because we feel that retrospective chart review is a sub-optimal approach to assessing causal influence unless the causal linkages between the processes and outcomes being examined are knowable (see page 419, column 2). We believe our estimates are the best estimates thus far based on implicit review and we haven't been able to come up with a superior, feasible approach to addressing this question, but that doesn't mean that we should pretend that we know the answer. The conclusion of our study is that if you accept chart review as a gold standard, than the often-quoted statistics are sensationalized over-estimates. From now on if people say the HMPS suggests that, "There are as many as 98,000 deaths in which two physician reviewers reported that an important error probably or definitely caused the death", than that quote may be misleading but it would be technically accurate. But no one should say that the HMPS suggests that, "There are as many as 98,000 people killed each year in US hospitals by medical errors". The HMPS cannot support that claim. The assumptions upon which the statistics are based have been shown to be incorrect in the only study that has rigorously evaluated them.

5) Some have suggested that results from observational studies evaluating risk-adjusted outcomes support much higher preventability numbers.

Some suggested that if you consider the variability in mortality rates between hospitals, or evidence from observational studies on volume-outcome associations that you come up with preventability numbers that are close to the IOM statistics. This raises a completely new issue and is beyond the scope of our paper, but we will try to briefly deal with this issue here anyway.

The statistics quoted in the IOM report and by the media come from chart review studies, not analyses of variations in risk-adjusted outcomes measures. Our study was designed to examine what the best estimate would be if chart review data is analyzed with consideration of 1) the attributable risk of the quality problem, 2) the skew in inter-rater reliability, and 3) short-term life-expectancy. Although we are very supportive of quality improvement interventions like increasing the expertise, skill and experience of providers (eg, increase use of intensivists and hospitalists, having a pharmacist round with the ICU team and decreasing the number of low volume facilities), as we state in the paper, we think that the current state of risk-adjustment precludes using non-experimental designs as clear evidence for the magnitude of the process-quality link. First, most of the large-scale risk-adjusted outcomes studies have used only mediocre case-mix adjusters, and second, even state-of the-art case-mix is not able to completely account for selection biases and endogeneity. Several studies have been unable to find any process differences between facilities with markedly different risk-adjusted mortality rates and even those studies that have identified process differences have not shown that the process differences account for a substantial proportion of the observed risk-adjusted mortality differences.

In addition, observational studies suggestion that transfer patients have much worse outcomes (including mortality rates that are up to 2 times higher) even after full risk-adjustment but I do not think many believe that the poorer outcomes are caused from the act

of transferring the patient. It's much more likely that unmeasured severity in transfer patients accounts for these poorer outcomes. Similarly, observational studies have demonstrated that people who volunteer for studies, take their placebo pills every day or receive academy awards10 have much lower mortality rates, but I do not believe that forcing people to do volunteer work, take placebo pills and giving everyone academy awards would result in everyone living much longer. There are often major self-selection, referral and social issues that determine who ends up at academic hospitals, in specialized stroke units or at high-volume surgical centers, and it's very difficult to determine how much of the observed risk-adjusted effect size is due to the exposure vs. unmeasured factors endogenous to the patient.

Still, although we suspect the numbers are exaggerated in this case also, we certainly believe there is substantial evidence that we can improve quality and outcomes by improving the quality and expertise of the personnel providing care. That conclusion is actually very consistent with the findings of our study, which suggest that major blunders are a rare cause of preventable death, but there are many questionable decisions made in complex, severely ill people in which having highly skilled and experienced professionals making those decisions may be the single most valuable quality improvement and patient safety intervention. Improving support systems is also important, but the types of problems found in chart reviews suggest that the most important factor is probably the quality of the providers using those support systems.

Although a few people have misinterpreted our paper, we have been very gratified by the overwhelmingly positive response that we've received, including by many who are very devoted to quality improvement and patient safety. Please do not hesitate to contact us if you have questions about the paper.

Rod Hayward, MD
Director, VA Center for Practice Management & Outcomes Research
Director, Quality Enhancement Research Initiative for Diabetes Mellitus (QUERI-DM)
VA Ann Arbor Healthcare System

Professor of Medicine & Public Health
Associate Director, RWJ Clinical Scholars Program
University of Michigan

Tim Hofer, MD, MSc
Senior Investigator, VA Center for Practice Management
& Outcomes Research
Co-Director, Patient Safety Research Program
VA Ann Arbor Healthcare System

Associate Professor of Internal Medicine
University of Michigan School of Medicine

References:

1. Brennan TA, Leape LL, Laird NM, Herbert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients.
Results of the Harvard Medical Practice Study I. N Engl J Med 1991 Feb;324(6):370-6.

2. Brennan TA. The Institute of Medicine report on medical errors--could it do harm? N Engl J Med. 2000 Apr 13;342(15):1123-5.

3. Leape LL. Institute of Medicine medical error figures are not exaggerated. JAMA. 2000 Jul 5;284(1):95-7.

4. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in Institute of Medicine report. JAMA. 2000 Jul 5;284(1):93-5.

5. Sox HC, Woloshin S. How many deaths are due to medical error? Getting the number right. Eff Clin Pract 2000;6:277-283.

6. Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. BMJ 2000 Feb 19;230(7233):469-473.

7. Christakis NA. Predicting patient survival before and after hospice enrollment. Hosp J 1998;13(1-2):71-87.

8. Addington-Hall JM, MacDonald LD, Anderson HR. Can the Spitzer quality of life index help to reduce prognostic uncertainty in terminal care? Br J Cancer 1990;62:695-9.

9. Evans C, McCarthy M. Prognostic uncertainty in terminal care: can the Karnofsky index help? Lancet 1985;148:1204-6.

10. Forster LE, Lynn J. Predicting life span for applicants to inpatient hospice. Arch Intern Med 1988;148:2540-3.

11. Hofer TP, Kerr EA, Hayward RA. What is an error? Eff Clin Pract 2000;3(6):1-10s.

10. Redelmeier DA, Singh SM. Survival in Academy Award-winning actors and actresses. Ann Intern Med 2001 May 15;134(10):955-62.

12. Localio AR, Weaver SL, Landis JR, Lawthers AG, Brenhan TA, Hebert L, Sharp TJ. Identifying adverse events caused by medical care: degree of physician agreement in a retrospective chart review. Ann Intern Med 1996 Sep 15;125(6):457-64.

13 Anderson RE. An "Epidemic" of Medical Malpractice? A Commentary on the Harvard Medical Practice Study. New York, NY: Manhattan Institute; 1996. Manhattan Institute Civil Justice Memo No. 27. Available at http://www.manhattan-institute.org/html/cjm_27.htm.

Methodological Appendix (.pdf file)

"Estimating Hospital Deaths Due to Medical Errors: Preventability Is in the Eye of the Reviewer" (.pdf file)

Rod Hayward & Tim Hofer

JAMA, July 25, 2001 - Vol 286, No. 4

The measurement model

The reviewers rated preventability on a 0-100 scale representing the probability of survival had care been optimal. Their ratings were highly skewed toward higher ratings of preventability, but normalized by a log odds transformation suggesting that the reviewers estimated odds of survival on a multiplicative scale. We constructed a hierarchical model for the log odds of survival that can be represented mathematically as follows:

$$Y_{ij} = \beta_0 + u_i + e_{Ij}$$

with u_i ~ iid $N(0,\tau_{00})$ and e_{ij}~ iid $N(0,\sigma^2)$

where:

Y_{ij} = the logodds of estimated survival with optimal care of the ith patient by the jth physician reviewer

β_0 = grand mean of Y (the log-odds of survival)

u_i = patient true log odds survival as deviations around the grand mean

e_{ij} = variation across the reviews within patient

τ_{00} and σ^2 are the variation in between-patient and between-review differences respectively where the
differences are independent and identically distributed (iid)

In the hierarchical model three parameters were estimated, the constant β_0, τ_{00}, *and* σ^2. A test for reviewer effect in a cross-classified hierarchical model did not show a significant reviewer effect. (See also Hofer et al for further details of these regression models examining the reliability of physician review.)[2]

Estimating the effect of unreliability and rating skew

It is well known that simply exponentiating log transformed model estimators (E(ln(Y))) will lead to a biased posterior estimate of Y [E(Y)]. Therefore the posterior or shrunken means for each patient or u_i's were calculated[1-3] and in a Monte-Carlo simulation 100 Y_{ij}'s per patient were generated by drawing from the estimated distributions of β_0 and e_{ij} keeping the u_i's fixed by patient. The Y_{ij}'s were then back transformed to the 0-100 probability scale by the inverse of the log-odds transformation. (This is an alternative method to the "smear" estimate[4] and this type of post-model-estimation simulation technique is described accessibly in King et al for more complex models.)[1] This allows us to examine the effect of using 100 reviews per patient, based on the measurement characteristics of implicit physician review, to estimate preventability and to take either the mean or the median of a simulated 100 reviewers.

1. King G, Tomz M, Wittenberg J. Making the most of statistical analyses: Improving interpretation and presentation. American Journal of Political Science. 2000 Apr; 44(2):341-355

2. Hofer, T. P.; Bernstein, S. J.; DeMonner, S., and Hayward, R. A. Discussion between reviewers does not improve reliability of peer review of hospital quality. Med Care. 2000 Feb; 38(2):152-61.

3. Goldstein, Harvey. Multilevel Statistical Models. 2nd ed. New York: Halstead Press; 1995.

4. Duan N, Manning WG. A comparison of alternative models for the demand for medical care. J Econ Bus Stat 1983;1:115-126.s